

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
16 December 2004 (16.12.2004)

PCT

(10) International Publication Number
WO 2004/109546 A1

(51) International Patent Classification⁷: **G06F 17/30**

(74) Agent: **PAGE WHITE & FARRER**; Runeberginkatu 5,
10th Floor, FI-00100 Helsinki (FI).

(21) International Application Number:
PCT/FI2004/000331

(22) International Filing Date: 1 June 2004 (01.06.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
20030855 6 June 2003 (06.06.2003) FI

(71) Applicant (for all designated States except US): **TIETOE-
NATOR OYJ** [FI/FI]; Kutojantie 10, FI-02630 Espoo (FI).

(72) Inventors; and

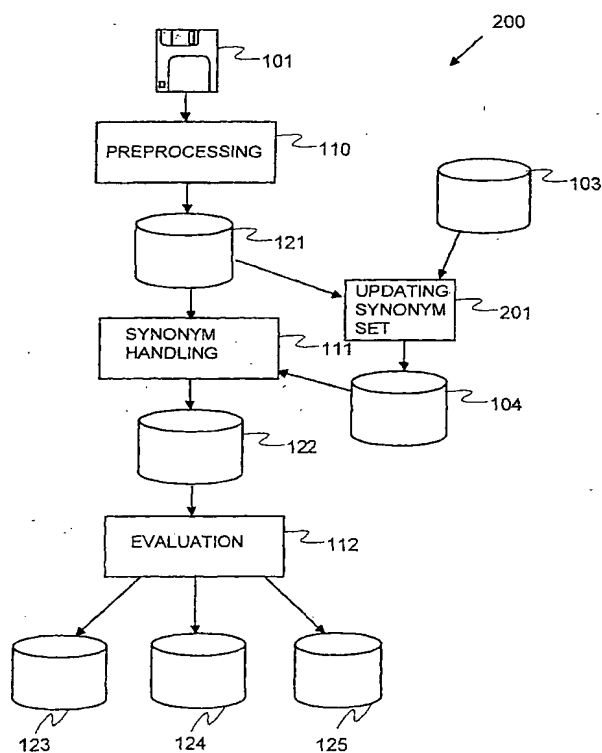
(75) Inventors/Applicants (for US only): **SURAKKA, Timo**
[FI/FI]; Länsituulentie 8 B 29, FI-02100 Espoo (FI).
PULKKINEN, Pentti [FI/FI]; Meiramikuja 3 B 6,
FI-02940 Espoo (FI).

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AL, AM,
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,
ZW.

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI,

[Continued on next page]

(54) Title: **PROCESSING DATA RECORDS FOR FINDING COUNTERPARTS IN A REFERENCE DATA SET**



(57) Abstract: The present invention relates to a method of processing a data record for finding a counterpart in a reference data set. In the data record a value of a data field is determined, where the data field represents an identifier. At least one synonym candidate for the value of the data field is determined from a set of predetermined identifier values. It is determined if a synonym candidate and the value of the data field fulfill a predetermined synonym acceptance criterion, and if the predetermined synonym acceptance criterion is fulfilled, the value of the data field and the synonym candidate are associated as synonyms. Thereafter a search for a counterpart for the data record is conducted by comparing to entries of the reference data set the value of the data field and/or a synonym associated with the value of the data field.



SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Published:

- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*